

## **DATA ANALYSIS PREPARATION AND BASIC ANALYSIS**

In the last lecture, ethnographies and surveys were presented and we discussed how researchers collect data and measure variables using survey questions. The next step for the researcher after collecting all survey data about a research question is interpreting what was collected. First, all raw data, or every value for every variable in the observation, is presented in a data matrix. A data matrix is two-dimensional table used to organize and retrieve raw data. In a data matrix, the numbers presented are not necessarily informative. For example, for a variable like gender, the number 1 might be presented in a data matrix which could mean male or female. Within a data matrix there are both columns and fields. A column in a data matrix is just one character (number or letter) wide; this is not the same as a column in a spreadsheet or printed table. Like in the previous gender example, only one column is needed because gender is demarcated with either a 1 or a 2. In contrast, a field is a series of adjacent columns used to represent a single variable. In this way, a field *is* like a column represented in a spreadsheet or table. Since data matrices include only raw data that include characters that are not meaningful, a codebook must be used. A codebook is a listing of all the variables in a data collection with each variable accompanied by a short variable name, a brief description, the location, and the meaning of each value. Without the codebook, the data matrix is meaningless.

### **Univariate Statistics and Bivariate Tables**

When samples of people are questioned (e.g., by using a survey), the amount of data that is collected can be quite large. Even if a researcher only asked one question to 200 people, it would be difficult if not impossible to make sense of all the data. Because it is impossible to eyeball a survey and interpret any meaning from it, researchers must use of different statistics.

(Don't worry, we will not go into statistics in great detail. We will simply provide enough information needed for you to be conversant in the topic.) One statistical method used is **univariate analysis**. Univariate analysis is used to examine summary statistics one variable at a time. Univariate analysis makes use of both measures of central tendency and measures of dispersion. Measures of central tendency include statistics such as the mean, median, and mode, and measures of dispersion include statistics that represent scatter such as the range and standard deviation. Univariate analysis can be conducted at the nominal, ordinal, interval, and ratio level. As a refresher:

<b>Level of Measurement</b>	<b>Statistics</b>
Nominal	Mode
Ordinal	Mode, Median
Interval/Ratio	Mode, Median, Mode, Range, Standard Deviation

A bivariate table shows the relationship between two variables. In a bivariate table, generally one of the variables is the independent variable (IV) and one is the dependent variable (DV). An example of a bivariate table is the following:

<b>Ever Been Hit By Anyone</b>	<b>Male</b>	<b>Female</b>	<b>Total</b>
<b>Yes</b>	53.1	26.2	38.2
<b>No</b>	46.9	73.8	61.8
<b>Total %</b>	100	100	100
<b>Total n</b>	228	282	510

In this table, the IV would be gender (i.e., being either male or female) and the DV is whether one has been hit or not. In this example it is easy to see how the IV would precede the DV (this is because it makes more sense that being male or female determines whether one has been hit than the opposite), but sometimes figuring out which is the IV and which is the DV can be difficult or even impossible. Reading bivariate tables seems easy, but can be trickier than you think. To better read tables follow these simple rules:

- If the table percentagizes down, compare across. Since the example table percentagizes down, then you compare across. This way, males can be compared to females.
- Generally, tables are percentagized within the categories of the independent variable. While this is usually the case, it is best practice to determine the independent variable based on what makes more logical sense. For example, does it make more sense that gender determines if you have been hit or whether you have been hit determines your gender?

## **Summary**

After the survey, questionnaire, interview, etc. is completed, the next step for researchers to do is pull the information collected and interpret it. In order to do this, researcher must create a data matrix and code the data. Afterwards, data can then be entered into specific tables such as a bivariate or trivariate table. Once all of this is completed, the data will be available for analysis which will ultimately allow for the preparation of a report. We will cover these topics in the upcoming weeks.

## **Weekly Assignment**

Your assignment for this week is as follows:

1. Read the notes for Week #11
2. Read Chapter 10 of your book
3. Post (on the Discussion Board) a brief overview of univariate statistics and bivariate tables. What type of information can be conveyed from these statistics and tables?